A Consumer-focused Modular Approach to Labeling IoT Devices and Software

Removed for review

August 2021

Abstract

Prior research has shown that nearly the majority of people do not read privacy policies and consequently, cannot make a fully informed privacy decision. Given the complex data ecosystem in Internet of Things (IoT) devices and software, privacy policies are more opaque and less accessible. Presidential Executive Order 14028 requires a robust "labeling program" for consumers to know how their data is being collected, used, and stored [10]. Our modular approach to consumer-focused privacy and security labels offers a model for such labels. It leverages existing Records of Processing (RoP) to form a robust, scalable, and platform-independent privacy label for consumers to use. We tested our approach on a complex social media privacy policy and an initial pilot study showed promising results - the labels were more usable and informative. This approach can be leveraged to both IoT devices and software labeling as we strive towards more consumer-driven privacy preservation.

Background and Problem Statement

An evaluation of behavior on the web found that 59% of internet users have never read a privacy policy [14], so they are unaware of the extent to which information is being collected about them and how it is being used. Even if consumers do read privacy policies, they opaque. Additionally, as Schaub notes in his work [19], privacy policies are often binary and ineffective. This is means that the consumers are forced to either agree to the privacy policy entirely or forego the possibility of using the particular software or IoT device related to the policy. As more and more sensitive consumer data is collected by organizations and existing privacy regulations (like the Health Information Portability and Accountability Act (HIPAA)) are limited in technological adaptation (e.g., HIPAA not applying to fitness trackers), it is increasingly important that consumers are able to make an informed decision about data compilation, use, and sharing.

In this position paper, we offer for consideration an approach for developing software and IoT privacy labels that are modular, with each service offered by the organization corresponding to a label tag. This is helpful to consumers who might choose to be connected to the organization as a consumer, but do not want their data to be used or shared for unwanted services. Such a mechanism would not only be much more granular, but also compatible with existing regulatory demands and the overall National Institute of Standards and Technology (NIST) Privacy Framework. A modular approach to privacy labels has the following advantages:

- It is aligned with rapid innovation. It enables addition, modification, and deletion of label tags based on services that are supported at any given time.
- It is practical and does not add additional criteria to existing Records of Processing (RoP). If organizations already have a data and processing activities map, they can simply export it to the label format.
- It is compatible with with current international privacy regulations as long as the RoP remains updated and accurate.
- Usable and acceptable for consumers.
- It can accommodate a plethora of software and IoT devices and services based.
- It can be adjusted to security guidelines and criteria as they become available. Once IoT cybersecurity and secure software development criteria are developed, they can be immediately incorporated into the modular privacy policy.
- It is flexible. It inherently enables allow manufacturers to adapt and built functionality into IoT devices based on consumer-specific demands for data protection.

Modular privacy policies can also encourage innovation in manufacturers' IoT security efforts, without building in assumptions about future technologies and the security landscape. The *feasibility and possible means for implementing tiered labels that reflect increasingly comprehensive levels of testing and assessment* is illustrated by prior work on privacy policies on Facebook [4].

A Modular Solution

Prior research has supported the use of task-based or session-based user-adaptive systems to make systems more trustworthy [12, 6]. Task-based systems inform users about *how* their data is being used and also that their data is not being used for purposes to which they have not consented. In the context of IoT devices, consumers of IoT devices should be able to restrict information shared for one purpose (e.g., turning on the coffee machine) to be used for a different purpose (e.g., receive advertisements for available coffee delivery services) unless the service is desired. In addition, data applied for a certain context should not be retained beyond that context (e.g., location data should be removed once weather information is provided to the requesting user). The modular privacy framework integrated Nissenbaum's 'Contextual Integrity' [16] and Camp's 'Design for Trust' [2] approaches to generate the preference-specific interface by considering the user (stakeholder), usage (purpose), and the environment variables. We also applied the additional guidelines provided in Cranor's 2004 [3]:

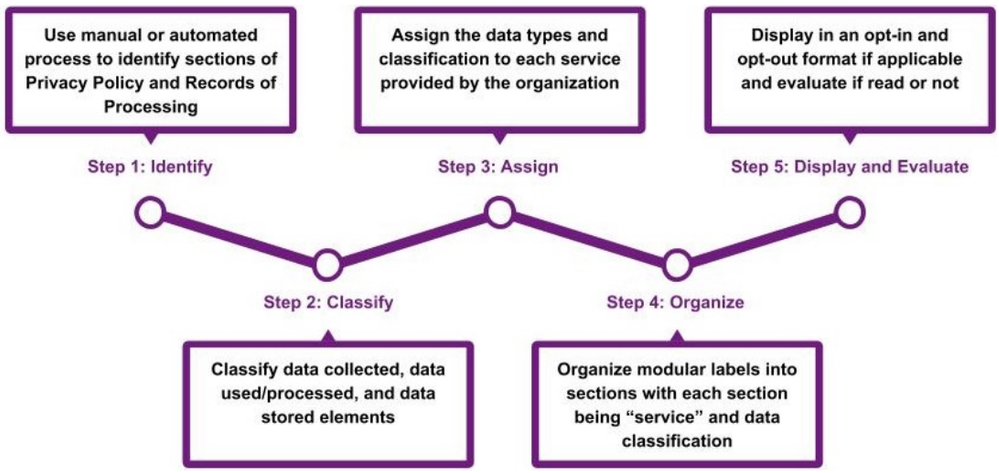

Figure 1: Step-by-Step Approach to Modular Labels

- Users must be able to update their preferences at any time. This is to return the *'preference control remote'* back to the user.
- The system must be open about *what* data is being collected, *how* it is being collected, and for *how long* it is being retained.
- The least possible user information should persist beyond session lifetime.
- Personalized interfaces must be built with sensitivity to the risks context. Thus, organizations which collect sensitive data such as financial services, health and disability services, differ from games. Services which either support potentially vulnerable populations or could be used to harm those targeted for harassment should meet a higher standard of care (e.g., [17])

Building on these principles we developed a pilot implementing a modular privacy policy labeling. The privacy policy is divided into sections, one for each of the respective services. The interaction enables consumers to agree to specific sections of a privacy policy. Since a privacy policy is required to state the data collected by each service, it was straight-forward to create these separate modules. This interaction enables consumers to clearly see what data is collected by each services. The modular label can also be extended to be more interactive so consumers can also decline to use a particular service by opting out of the privacy policy module corresponding to it. Each module identifies a service, the data, and the recipients of the data. The same consumer may feel very differently about the same datum based on reason for use and recipient [8, 1].

We create a modular privacy policy model by implementing the following design iterations: (i) Study the current data policy either manually or through automation, (ii) Develop sections for classification (which services are being offered) derived from the data policy, (iii) Extract sections from policy and fit into the classification of service to policy mapping (this mapping contains information on data collection, usage/processing, and storage), (iv) Render the modular sections of the policy based on the extracted data in a printed label format, and v) Validate the usability of the modular policy by conducting qualitative user studies. Our initial pilot study suggested that participants were more willing to pay attention to the labels and they understood what data was being processed and how it was being used. At the user end, these service-specific policy statements can be seen and accepted in parts according to the particular service they desire. [4].

In the current label used in the pilot, there are two core design goals: (i) Develop a conceptual model for selective privacy policy statements. (ii) Improve clarity and visibility in the policy model through interface design.

Generic user modeling systems provide a basic framework for studying how users value their privacy. In networked systems and services, it is becoming increasingly complex to understand and resolve conflicting privacy values and usability issues [11, 8]. Machine Learning could also be used to automate the generation of labels based on their use in privacy policies [20, 9, 13, 15, 21]. One of the machine learning approaches that directly informed our work was the framework proposed by Harkous et al. [9]. Their proposed framework uses a hierarchy of neural network classifiers to identify high-level and fine-grained data collection and usage policy details. They demonstrate the utility of their framework by building an application that given a link to the privacy policy could automatically generate disconnect icons [1] based on preset rules. Using the same framework a similar application can be built to generate aggregate privacy ratings for privacy policies based on preset rules or user preferences.

The user modeling in the pilot was influenced by the work of Preibusch and Soren, who proposed a framework for 'analyzing privacy requirements' and matching them with the data collected by online services to ensure that no additional data is left unprotected [18]. The user modeling and interaction design was also informed by research on privacy perceptions and behaviors [7**?**]. Given the current international requirements and available tool, this approach to providing the required information to inform consumers should not significantly raise the cost on companies, since they already have this information on their data processing.

Application and Broader Impact

Our modular privacy label approach provides an approach which empowers consumers to take control of their privacy by selectively agreeing to only the sections of the privacy policy for the services they desire. These sections correspond to specific services offered by the platform or device which a consumer might or might not choose to use.

The modular approach provides transparency about and control over the specific information shared with manufacturers, service providers, and third-parties. Transparency would ensure that organizations not only remain compliant with existing regulations and the overall NIST Privacy Framework, but also increase trust through effective data

[1]Disconnect icons are privacy icons that were developed a Mozilla led working group. The aim of these icon was to make it easy for people to understand the terms of the privacy policy and to communicate data collection and usage practices [5].

communication. Additionally, we also believe that developers will find it more convenient to implement sections as privacy labels when they develop or upgrade new services without modifying an entire label. Finally, our proposed solution leverages organizations' existing RoP, which would reduce the burden on organizations to become compliant with a new labeling requirement.

Such an approach could be effective in providing consumers an opportunity to select different products and services based on the desired services; and enable them to purchase only that which they choose. However, the presentation of the options, including the timing and framing, are critical design decisions; for example, should consumers opt-out or opt-in? A design beyond a pilot must address issues of timing, human decision-making biases, and risk perception.

References

[1] Laura Calloway, Hilda Hadan, Shakthidhar Gopavaram, Shrirang Mare, and L. Jean Camp. Privacy in Crisis: Participants' Privacy Preferences for Health and Marketing Data during a Pandemic. In *Proceedings of the 19th Workshop on Privacy in the Electronic Society*, WPES'20, page 181–189, New York, NY, USA, 2020. Association for Computing Machinery.

[2] L Jean Camp. Design for trust in ambient and ubiquitous computing. In *International Conference on Autonomic and Trusted Computing*, pages 1–1. Springer, 2009.

[3] Lorrie Faith Cranor. I didn't buy it for myself. In *Designing personalized user experiences in eCommerce*, pages 57–73. Springer, 2004.

[4] Sanchari Das, Jayati Dev, and Kaushik Srinivasan. Modularity is the key a new approach to social media privacy policies. In *Proceedings of the 7th Mexican Conference on Human-Computer Interaction*, page 13. ACM, 2018.

[5] Disconnect. Disconnect privacy icons. https://web.archive.org/web/20170709022651/disconnect.me/icons. [Online; accessed 28-June-2019].

[6] Steve Dodier-Lazaro, Ruba Abu-Salma, Ingolf Becker, and M Angela Sasse. From paternalistic to user-centred security: Putting users first with value-sensitive design. In *CHI 2017 Workshop on Values in Computing*. Values In Computing., 2017.

[7] Vaibhav Garg, Kevin Benton, and L Jean Camp. The privacy paradox: a facebook case study. In *Telecommunications Policy Research Conference*, 2014.

[8] Hilda Hadan, Laura Calloway, Shakthidhar Gopavaram, Shrirang Mare, and L Jean Camp. American Privacy Perceptions in the COVID Pandemic. *Annals of Disaster Risk Sciences*, 3(2), 2020.

[9] Hamza Harkous, Kassem Fawaz, Rémi Lebret, Florian Schaub, Kang G. Shin, and Karl Aberer. Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning. In *27th USENIX Security Symposium (USENIX Security 18)*, pages 531–548, Baltimore, MD, 2018. USENIX Association.

[10] The White House. Executive Order 14028 on Improving the Nation's Cybersecurity, 2021.

[11] Alfred Kobsa. Generic user modeling systems. *User modeling and user-adapted interaction*, 11(1):49–63, 2001.

[12] Alfred Kobsa and Jörg Schreck. Privacy through pseudonymity in user-adaptive systems. *ACM Transactions on Internet Technology (TOIT)*, 3(2):149–183, 2003.

[13] Frederick Liu, Shomir Wilson, Peter Story, Sebastian Zimmeck, and Norman Sadeh. Towards Automatic Classification of Privacy Policy Text. 12 2017.

[14] Aleecia M McDonald and Lorrie Faith Cranor. The cost of reading privacy policies. *Isjlp*, 4:543, 2008.

[15] Abhijith Athreya Mysore Gopinath, Shomir Wilson, and Norman Sadeh. Supervised and Unsupervised Methods for Robust Separation of Section Titles and Prose Text in Web Documents. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 850–855, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.

[16] Helen Nissenbaum. Privacy as contextual integrity. *Wash. L. Rev.*, 79:119, 2004.

[17] Olivia Nuzzi. What it's like to get doxed for taking a bike ride, Jun 2020.

[18] Sören Preibusch, Bettina Hoser, Seda Gürses, and Bettina Berendt. Ubiquitous social networks: Opportunities and challenges for privacy-aware user modelling. 2007.

[19] Florian Schaub, Rebecca Balebako, and Lorrie Faith Cranor. Designing effective privacy notices and controls. *IEEE Internet Computing*, 2017.

[20] Shomir Wilson, Florian Schaub, Aswarth Abhilash Dara, Frederick Liu, Sushain Cherivirala, Pedro Giovanni Leon, Mads Schaarup Andersen, Sebastian Zimmeck, Kanthashree Mysore Sathyendra, N Cameron Russell, et al. The Creation and Analysis of a Website Privacy Policy Corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1330–1340, 2016.

[21] Sebastian Zimmeck and Steven M. Bellovin. Privee: An Architecture for Automatically Analyzing Web Privacy Policies. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 1–16, San Diego, CA, 2014. USENIX Association.